



Acta Otorrinolaringológica Española

www.elsevier.es/otorrino



ORIGINAL

Diseño de una nueva herramienta para la exploración auditiva

Antonio Rodríguez Valiente^{a,*}, Carolina Pérez Sanz^b, Carmen Górriz^a,
Adoración Juárez^c, Marc Monfort^d, José R. García Berrocal^a,
Juana Gil Fernández^e y Rafael Ramírez Camacho^a

^aServicio de Otorrinolaringología, Hospital Universitario Puerta de Hierro, Madrid, España

^bGabinete de Logopedia y Foniatría Frasis-Voz y Lenguaje, Madrid, España

^cColegio Tres Olivos, Madrid, España

^dCentro Entender y Hablar, Madrid, España

^eLaboratorio de Fonética, Consejo Superior de Investigaciones Científicas, Madrid, España

Recibido el 26 de marzo de 2008; aceptado el 10 de septiembre de 2008

PALABRAS CLAVE

Audiometría verbal;
Espectrograma;
Análisis acústico

Resumen

Introducción: muchos pacientes presbiacúsicos presentan dificultad para entender ciertas palabras, hecho que podría estar justificado porque determinados sonidos en español presentan una dificultad mayor de perceptibilidad, en concreto los de energía en altas frecuencias. Se propone utilizar una oración como herramienta para comprobar la hipótesis de partida.

Material y método: se analizaron todos los sonidos del español, midiendo el grado de energía acústica que presenta cada una de las frecuencias. Las conclusiones extraídas de la comparación de sus resultados permitieron el diseño de la herramienta que aquí se propone.

Resultados: se establece una gradación de dificultad perceptiva, por lo que se puede decir que las oclusivas aparecen como los sonidos menos perceptibles, seguido de las fricativas y, finalmente, todos los segmentos que presentan armonicidad y una estructura formántica definida. Como herramienta para la práctica clínica, se propone la frase española "Ana vio ese coche rojo fino", frase que cumple una serie de particularidades que la hacen especialmente útil para tal fin. Esta oración permitirá comprobar si la comprensión se va deteriorando a medida que se avanza desde la primera a la última de sus secciones, y así poder suponer que la capacidad para detectar auditivamente la presencia de energía reforzada en las altas frecuencias es indispensable para la inteligibilidad.

Conclusiones: un resultado positivo podría tener como consecuencia el diseño de sistemas de amplificación que mejoren la inteligibilidad de la palabra. Además, el hecho de tener una herramienta exploratoria podría permitir la exploración neuroacústica, de utilidad en el estudio de la enfermedad auditiva.

© 2008 Elsevier España, S.L. Todos los derechos reservados.

*Autor para correspondencia.

Correo electrónico: arova777@hotmail.com (A. Rodríguez Valiente).

KEYWORDS

Speech audiometry;
Spectrogram;
Acoustic analysis

Designing a new tool for hearing exploration**Abstract**

Introduction: Many presbycusic patients have difficulty in understanding certain words. This could be justified because certain sounds in Spanish are more difficult to perceive, particularly the sounds with energy in the high frequencies. We propose to use a sentence as a tool to check this theory.

Materials and method: All the Spanish sounds were analyzed, measuring the degree of acoustic energy in all the frequencies. The conclusions drawn from the comparison of the results allowed the design of the tool that is proposed here.

Results: We established a gradient of perception difficulty, occlusive consonants being the least perceptible, followed by fricative, and finally all those segments with harmony and a clear formant structure. The Spanish sentence "*Ana vio ese coche rojo fino*" is proposed as the tool for this study. This sentence has some as it comprises certain peculiarities that makes it particularly useful for this purpose. It will allow us to check whether understanding deteriorates as we move from beginning to end, helping evaluate the importance of high frequencies for intelligibility.

Conclusions: A positive result could help in the design of amplification systems to improve speech intelligibility. In addition, the exploratory tool could allow neuro-acoustic exploration, useful in the central auditory pathology studies.

© 2008 Elsevier España, S.L. All rights reserved.

Introducción

La mayor parte de los pacientes presbiacúsicos tienen dificultad para percibir y entender determinadas palabras integradas por sonidos concretos. Esta constatación aconseja diseñar una herramienta de investigación aplicada que resulte válida para comprobar la hipótesis de partida, a saber, que no todos los sonidos del español presentan un grado de inteligibilidad similar, y que esa diferencia de perceptibilidad entre ellos se fundamenta en la ubicación que presentan sus componentes en el espacio acústico disponible. Más en concreto, se quiere probar que la existencia en ciertos elementos fónicos de energía en muy altas frecuencias, con el condicionamiento en el rango de audición que ello supone, puede ser un factor decisivo en su grado de audibilidad.

Estas observaciones realizadas en la práctica clínica diaria han motivado esta investigación, en la que se propone utilizar una oración breve, justificada por su composición segmental y suprasegmental, con el fin de responder la demanda de recursos para la investigación propuesta.

Antecedentes del problema

La inteligibilidad disminuye progresivamente a lo largo de la vida, hasta el punto de que la presbiacusia muestra una disminución de la diferenciación, que aumenta con el tiempo. Se han diseñado diversas soluciones amplificadoras para resolver este problema. Pese a ello, un número determinado de pacientes aún se queja de mala inteligibilidad en situaciones adversas (ambiente ruidoso, varias personas hablando al mismo tiempo, etc.). Dado que la mayor parte de las prótesis cubren un rango de frecuencia que no supera los 4.000-6.000 Hz¹, estarían sin cubrir las frecuencias más afectadas por la presbiacusia.

Propuesta de herramienta

El español, y en concreto su variante castellana, es una lengua que favorece la inteligibilidad de los enunciados, ya que su sistema fonológico está integrado por un número reducido de fonemas que presentan relativamente pocos alófonos, y su subsistema vocálico es también limitado². Asimismo, en ella la proporción de palabras monosílabas es pequeña. Todo esto favorece la comprensión de los mensajes por parte del oyente, como demuestra el mejor rendimiento de adultos y niños hispanohablantes tras una implantación coclear, en comparación con el rendimiento de hablantes de otras lenguas. Con todo, la inteligibilidad de las emisiones no sólo depende de la lengua en la que se pronuncian, sino también de una serie de factores, entre los que pueden mencionarse la calidad de la dicción del emisor, la cualidad de su voz, las condiciones funcionales auditivas del receptor y el medio en el que se desarrolla el acto de comunicación. Estas variables pueden afectar en sentido positivo o negativo la transmisión de la información.

Para tratar de verificar la hipótesis, que tiene que ver, en concreto, con la influencia de la constitución acústica de los sonidos en su audibilidad, se ha optado por diseñar una oración en cada una de cuyas partes se concentrara un tipo determinado de sonido, en lo que respecta a la distribución frecuencial.

Como se sabe, ya se han confeccionado pruebas para diferenciar los trastornos que afectan la percepción fluida de enunciados orales³⁻⁵. Así, en la audiometría verbal, para valorar los diversos grados y tipos de pérdida auditiva, se emplean listas de palabras del español bisílabas y trisílabas. Su ventaja radica en que el sujeto explorado ha de identificar los componentes de las palabras emitidas y las ha de repetir en un entorno en el que se pierden las relaciones entre cada palabra y la siguiente, de forma que es imposible deducirlas por su significado o su

contexto. En este sentido, cabe señalar que la decisión de trabajar con un enunciado superior a la palabra se debe a la voluntad de crear un estímulo lo más cercano posible a la comunicación real, es decir, más complejo.

Material y método

En primer lugar, se procedió a grabar todos los sonidos del español pronunciados por un hablante masculino y otro femenino, tanto vocales como consonantes, aislados en el caso de las vocales o en posición fonéticamente normal en el de las consonantes, esto es, en el contexto *a_a*, para poder estudiar sus características intrínsecas, sin la influencia de la coarticulación con otros elementos. Era necesario hacerlo así también para que los sonidos tuvieran aproximadamente la misma duración, tono y acento, y de este modo su intensidad o energía acústica fuera comparable. La grabación se llevó a cabo en el Laboratorio de Fonética del Consejo Superior de Investigaciones Científicas, mediante el empleo del CSL 4500 de Kay Electronics Co, con una frecuencia de muestreo de 44.100 Hz. Se utilizó un micrófono de diadema modelo AKG C444L situado a 10 cm de la comisura labial del hablante y con un desplazamiento lateral de 45°, como marcan las directrices de procedimiento para el análisis acústico⁶⁻⁸.

A continuación, se procedió a editar las grabaciones, y se delimitaron los sonidos registrados, de manera que en ninguno de ellos, particularmente en las fricativas y oclusivas sordas, se percibieran fragmentos correspondientes a las transiciones con los sonidos vocálicos adyacentes. Una vez realizada esta tarea, los sonidos ya editados se analizaron mediante el software PRAAT, de libre acceso, cuya herramienta de análisis Spectrum/Spectrum Tier permite comprobar cuál es el grado de energía acústica, medida en dB, que presenta cada una de las frecuencias componentes de cada uno de los segmentos. En la tabla 1, a modo de ejemplo, se reproducen los datos obtenidos en relación con las fricativas sordas.

Finalmente, la comparación de estos datos, esto es, el cotejo de las frecuencias de cada elemento y de su intensidad correspondiente o *sonía*, tanto en la voz masculina como en la femenina, permitió extraer algunas conclusiones que resultaron de utilidad para llevar a cabo el diseño de la herramienta que aquí se propone.

Resultados

Con los valores recogidos en el análisis, se deduce lo siguiente:

- Entre las consonantes fricativas, las llamadas *sibilantes* —que son las que se articulan con una abertura horizontal más estrecha que la abertura vertical (se denominan a veces “abombadas”), y que en español son la [s] y la segunda fase de [ʃ] (la “ch” de *hacha*)— ofrecen un nivel de energía considerable (de 46-55 dB) en la gama frecuencial que se mueve en torno a los 6.500-7.000 Hz. En cambio, las frecuencias bajas no aparecen reforzadas.
- Las fricativas no sibilantes, [f], [x] (la “j” de *baja*) y [θ] (la “z” de *baza* y la “c” de *cena*), sí muestran refuerzo

Tabla 1 Valores en dB de cada una de las 10 frecuencias más reforzadas en las consonantes fricativas [ʃ], [s], [x], [θ] y [f]. Las frecuencias más reforzadas en [ʃ] y [s] se encuentran en torno a los 6.500-7.000 Hz, mientras que en [x], [θ] y [f] están en la parte más baja del espectro. La diferencia fundamental entre [θ] y [f] se origina a partir de las últimas 3 frecuencias mostradas aquí: mientras que en [θ], la séptima frecuencia que aparece reforzada es 3.006 Hz, mientras que en [f] es 11.534 Hz

ʃ	Frec (Hz)	Ds (dB/Hz)
	7.034,7184	49,5939867
	6.721,24918	47,4102637
	6.121,66661	46,7913527
	6.014,04025	46,4455567
	6.793,98022	46,1998035
	7.129,76212	46,1967457
	7.100,41158	46,1375997
	6.591,87264	44,7069017
	7.846,0364	44,4583153
s	Frec (Hz)	Ds (dB/Hz)
	6.704,75938	55,5915904
	6.623,17329	52,7830834
	7.005,89096	52,4822078
	7.134,17278	52,2950957
	6.547,17732	51,0174019
	7.022,50095	50,574668
	6.990,30341	50,4960023
	7.100,02369	50,4922775
	6.783,13793	50,2037553
x	Frec (Hz)	Ds (dB/Hz)
	70,8179343	66,5173011
	137,942909	55,5470978
	113,134542	54,9647791
	1.755,31159	51,2579729
	1.627,88981	50,6611579
	211,476575	49,3627643
	1.655,72318	49,1738561
	33,0424905	48,390016
	1.690,64422	48,1926085
θ	Frec (Hz)	Ds (dB/Hz)
	53,3065217	57,9471076
	40,2614916	54,1348469
	4,66598407	51,5578407
	14,2150504	50,9677095
	86,8198371	49,4484912
	113,831054	48,4791122
	3.006,74241	48,2413807
	2.952,19691	47,5413852
	5.116,96267	46,6211333
f	Frec (Hz)	Ds (dB/Hz)
	21,8909035	59,9376243
	12,0531868	54,5582861
	53,4217708	51,2787228
	68,7385728	50,6013289
	39,3248854	50,2418703
	82,3634332	47,8169694
	11.534,6935	47,6140699
	14.392,2909	47,1712648
	11.621,3386	46,79903

Ds: densidad sonora; Frec: frecuencia.

Tabla 2 Valores en dB de cada una de las 10 frecuencias más reforzadas en las vocales [a], [e], [i], [o], [u]

a	Frec (Hz)	Ds (dB/Hz)
	937,656286	64,2417931
	1.125,55491	62,4495366
	749,586463	62,0523815
	186,967847	61,7017315
	1313,6144	60,6849474
	927,142618	60,4748342
	1.114,85166	58,9390213
	739,977936	57,765972
	373,933471	57,6877286
e	Frec (Hz)	Ds (dB/Hz)
	196,956074	66,3890339
	396,146016	66,3802251
	595,21857	65,946384
	188,012273	63,7817405
	383,928014	62,3257725
	581,736411	62,2029421
	573,176941	61,6391833
	566,180368	60,5731901
	376,776807	59,4434761
i	Frec (Hz)	Ds (dB/Hz)
	205,609909	74,7025744
	196,712741	69,0871327
	191,456775	66,0266579
	411,995353	65,9668429
	186,325534	62,574912
	402,200341	62,078027
	395,208526	58,9933482
	180,797274	58,6331557
	387,992402	55,9892634
o	Frec (Hz)	Ds (dB/Hz)
	573,248305	66,1074428
	381,501714	64,7810525
	188,791285	63,9183019
	561,237597	61,9475573
	552,720448	60,2597225
	371,109125	60,0962881
	546,849893	58,0828707
	957,093812	57,9745194
	765,287739	57,7010426
u	Frec (Hz)	Ds (dB/Hz)
	200,14247	71,7034853
	401,777924	70,7596058
	190,751281	66,752235
	390,709272	66,6172194
	383,056985	63,8576324
	376,370384	63,3542002
	371,610492	62,6704557
	366,557911	61,4758289
	361,614772	59,260932

Ds: densidad sonora; Frec: frecuencia.

en frecuencias bajas, en particular la [θ] (p. ej., 54 dB en 40 Hz) y la [f] (p. ej., 51 dB en 53 Hz). En el caso de esta última, el refuerzo de la energía sonora es apenas perceptible entre los 80 Hz y la gama comprendida entre los 11.500 y los 14.800, mientras que en la [θ], a partir de los 2.800 Hz la energía aparece más uniformemente distribuida hasta la zona de frecuencias más altas: 41 dB en los 11.300 Hz.

En consecuencia, por lo que a las fricativas se refiere, y partiendo de la hipótesis inicial de que los sonidos con frecuencias más altas requieren un rango de audición mayor, o, si se prefiere, entrañan una dificultad mayor para pacientes con presbiacusia, podría establecerse una escala de mayor a menor grado de perceptibilidad como la siguiente:

[x] [θ] [tʃ] [s] [f]
+ _____ -

- c) En cuanto a las vocales, en primer lugar, conviene señalar que son sus primeros 2 formantes (F1 y F2) los que determinan su timbre, esto es, su identidad como sonido: una [a] como sonido diferente de una [e], de una [i], etc., mientras que los restantes formantes superiores (F3, F4 y F5) reflejan las características individuales de cada voz. Los datos obtenidos apoyan la idea de que todas ellas son sonidos fácilmente perceptibles, puesto que todos sus formantes se sitúan entre 190 y 6.000 Hz, con la zona más decisiva para su identificación situada en torno a los valores que se recogen en la tabla 2.
- d) El análisis de las oclusivas ([p], [t], [k]) demuestra, en primer lugar, que no presentan altas frecuencias dotadas de una densidad sonora significativa, y, en segundo lugar, que, como se ha apuntado repetidas veces en la bibliografía, son sonidos intrínsecamente menos perceptibles, puesto que tienen menor sonía en general: se producen por un cierre u oclusión en alguna zona del tracto vocal, que se rompe con un ruido característico denominado explosión, por lo demás no siempre presente. Además, el grado máximo de intensidad que alcanzan nunca llega a ser tan marcado como el propio de otros sonidos. De todo lo anterior, se deduce que su difícil percepción por los oyentes, incluso por aquéllos sin enfermedad alguna, no viene dada por la existencia de altas frecuencias que puedan quedar fuera de su gama auditiva, sino por su propia naturaleza.

Es interesante que, en la tradición de estudios lingüísticos de carácter fonético, se hayan propuesto una y otra vez durante años "escalas de sonoridad" o "sonía"^{9,10}, que ordenan los distintos tipos de sonidos (fig. 1).

Como puede observarse, estas escalas coinciden a grandes rasgos con la gradación de dificultad perceptiva deducible de lo hasta ahora visto aquí, es decir, las oclusivas aparecen como los sonidos menos perceptibles por lo explicado arriba, en el punto d); a continuación, son las fricativas las menos "sonoras" (entendiendo "sonoras" no como producidas mediante la vibración de las cuerdas vocales, sino como dotadas de "sonía"), y, finalmente, todos los segmentos que presentan una cierta armonicidad y una estructura formántica definida, como las nasales, laterales, vibrantes y las vocales, resultan ser los más perceptibles.

En efecto, esto se ha señalado repetidas veces: los sonidos que manifiestan una estructura formántica prominente y poco amortiguada son los más “resonantes”. La resonancia en el habla se ve siempre subrayada por la sonoridad, de modo que las cavidades resonadoras que integran el tracto vocal responden de una forma más eficiente cuando hay vibración en las cuerdas vocales, en cuyo caso la concentración de energía acústica en bandas de frecuencia determinadas es mayor y mucho más evidente. En este sentido, podría afirmarse que el nivel de perceptibilidad o sonía no depende únicamente de la intensidad del sonido, es decir, del grado de energía que presenta, sino de cómo se distribuye esa energía en su espectro¹¹.

Propuesta y justificación de la frase modelo

A partir de las constataciones a las que se llegó, después de los análisis descritos previamente, se ha considerado que un enunciado que resulta válido como herramienta para la práctica clínica es la frase española “Ana vio ese coche rojizo fino”, cuyo espectrograma se reproduce en la figura 2. A continuación se explica por qué se ha llegado a esa conclusión:

- La oración propuesta es una frase afirmativa simple, integrada por sujeto, verbo y complemento, que desde el punto de vista suprasegmental (es decir, por lo que se refiere a la melodía, la acentuación y el ritmo con el que se emite) no presenta particularidades destacables que puedan ayudar al oyente en el proceso de percepción de los sonidos. Por ejemplo, se sabe que una frase interrogativa favorece sistemáticamente la percepción de las frecuencias agudas, que se ven reforzadas en algunas posiciones; por ello se ha elegido un enunciado con una curva entonativa asertiva, digamos, neutra.
- La frase tiene una longitud reducida y constituye un único grupo fónico, esto es, un fragmento de enunciado susceptible de ser emitido entre 2 pausas, por ejemplo inspiratorias, sin interrupciones internas. El número de sílabas que lo componen responde a la media establecida para el español.
- El significado de las palabras de la secuencia no ayuda a entender su significado. No es predecible el adjetivo

“rojizo” a partir del sustantivo “coche”, mientras que si lo sería en mayor grado, por ejemplo, a partir de “cabello” o “pelo”.

- En la oración diseñada se pueden distinguir 3 partes que difieren entre sí por el tipo de sonidos que conllevan. “Ana vio” constituye el primer fragmento, y en él todos los elementos que se suceden (incluida la labial fricativa sonora “v” [β]) tienen sonoridad, tienen estructura formántica clara o, lo que es igual, son resonantes en el sentido que arriba se explicaba, presentan energía en diversas zonas de su espectro, y por ende son sumamente perceptibles; “ese coche” es la segunda parte de la frase, y en ella el grado de dificultad para la audición se acrecienta, porque, aunque hay vocales medias ([e] y [o]) que, en principio, son bastante perceptibles, aparecen 2 sibilantes, la [s] de “ese” y la segunda fase de la [ʃ], “ch” de coche, que, como se ha visto, son fricativas con gran cantidad de energía desplegada en las áreas frecuenciales altas. Finalmente, “rojizo y fino”, la última sección del enunciado, es la más complicada desde la perspectiva de la audición, dado que presenta varias vocales agudas [i], y los 2 sonidos más difícilmente perceptibles, como se ha visto, la [θ], “z” y la [f], además de la [x], “j”.

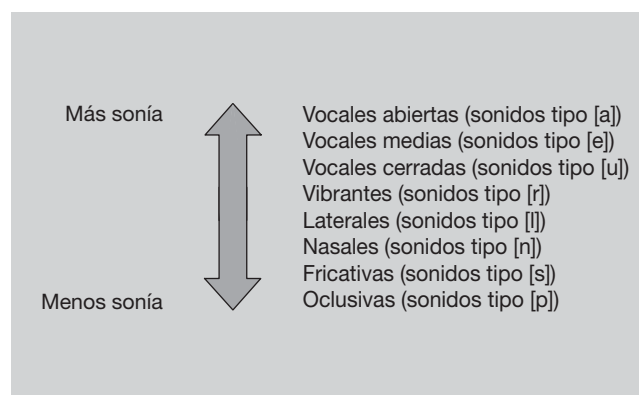


Figura 1 Escala de sonoridad que ordena los distintos tipos de sonidos.

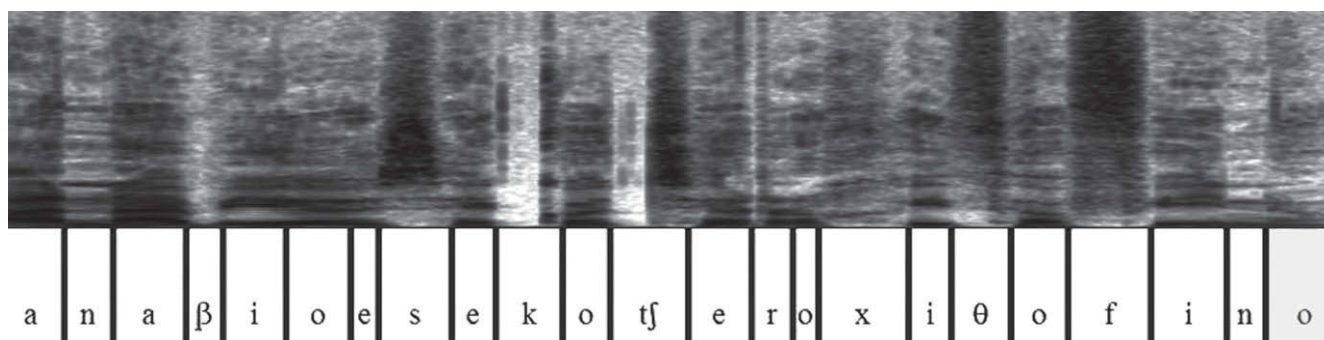


Figura 2 Espectrograma de banda estrecha de la frase propuesta “Ana vio ese coche rojizo fino” —grabada con una frecuencia de muestreo de 44.100 Hz—, obtenido con Praat, una frecuencia máxima de 22.050 Hz, dibujado en una ventana Hanning de 2 ms. Obsérvese como en el sonido [θ] la distribución de la energía sonora es más homogénea en todas las frecuencias, mientras que en [f] la zona de mayor ruido se concentra en la zona más alta, y en [s] y en la fase fricativa de [ʃ] se localiza en la zona media del espectro. En todas las vocales se distinguen armónicos hasta frecuencias de alrededor de 12 kHz, si bien [i] los presenta también más elevados; sin embargo, la mayor carga sonora en todas ellas se encuentra en los primeros formantes.

Si se comprueba, en la experimentación clínica, que la audición y consecuente comprensión de esta oración se va deteriorando a medida que se avanza desde la primera a la última de sus secciones, entonces habrá base suficiente para suponer que la capacidad para detectar auditivamente la presencia de energía reforzada en las altas o muy altas frecuencias es indispensable para la inteligibilidad de determinados mensajes, y se hará más evidente por qué determinados pacientes que ya no tienen esta capacidad presentan dificultades especiales para conseguir su comprensión.

Conclusiones

La posibilidad de asociar la capacidad de percepción por parte del oído afectado de una sordera neurosensorial de una oración diseñada con criterios fonéticos ("Ana vio ese coche rojizo fino") supone el desarrollo de una herramienta finalista que puede tener, como resultado, el diseño de sistemas de amplificación que mejoren la inteligibilidad de la palabra. La frase propuesta puede administrarla al oído explorado la propia persona mediante dispositivos electrónicos que contemplen sexo, edad, ruido ambiente, coincidencia con otras voces, que permitan la valoración de la percepción auditiva en condiciones próximas a la realidad. Adicionalmente, su asociación a técnicas ya implantadas, como la logaudiometría, podría permitir la exploración neuroacústica de áreas asociativas frente a áreas auditivas primarias, de posible utilidad en el estudio de la enfermedad auditiva central.

Bibliografía

1. Moore BC, Stone MA, Alcántara JI. Comparison of the electroacoustic characteristics of five hearing aids. *Br J Audiol.* 2001;35:307-25.
2. Maddieson I, Precoda K. UCLA Phonological Segment Inventory Database. (Base de datos de 451 lenguas disponible para descarga.) Disponible en: www.linguistics.ucla.edu/faciliti/sales/software.htm
3. Tato JM, Alfaro A. Audiometría del lenguaje. Discriminación por vía ósea: umbral de captación. Índice de utilidad social auditivo. *Otorrinolaringológica.* 1949;1:3.
4. Poch Viñals R. La exploración funcional auditiva. Madrid: Ed. Paz Montalvo; 1958.
5. Cárdenas MR, Marrero V. Cuadernos de Logaudiometría. Cuadernos de la Universidad Nacional de Educación a Distancia. Valladolid: Simancas Ediciones; 1994.
6. Deliyski DD, Evans MK, Shaw HS. Influence of Data Acquisition Environment on Accuracy of Acoustic Voice Quality Measurements. *J Voice.* 2005;19:176-86.
7. Titze IR. Toward standards in acoustic analysis of voice. *J Voice.* 1994;8:1-7.
8. Titze IR. Workshop on Acoustic Voice Analysis: Summary Statement. National Center for Voice and Speech; Denver, Colorado: 1994.
9. Clements GN. The role of sonority in core syllabification. En: Kingston J, Beckman ME, editors. *Papers in laboratory phonology I. Between the grammar and physics of speech.* Cambridge: CUP; 1990. p. 283-333.
10. Butt M. Sonority and the explanation of syllable structure. *Linguistische Berichte.* 1992;137:45-67.
11. Clements GN. Does sonority have a phonetic basis? *Laboratoire de Phonétique et Phonologie. CNRS / Sorbonne-Nouvelle, Paris, France* 2006. Disponible en: http://ed268.univ-paris3.fr/lpp/publications/2006_Clements_Does_Sonority.pdf